

WHAT IS CLAIMED IS:

1 1. A method of providing a single system image in a clustered environment
2 comprising:

3 assigning an internet protocol (IP) address as a cluster IP address;

4 binding the cluster IP address to a node in a cluster;

5 receiving a client request directed to the cluster IP address;

6 multicasting the request to all nodes in the cluster;

7 filtering the request based on a dynamically adjustable workload distribution

8 function, wherein the function is configured to allow a single node to process the client
9 request;

10 obtaining a response to the request;

11 inserting a cluster media access control (MAC) address into the response;

12 sending the response from the single node to the client.

1 2. The method of claim 1 further comprising informing other nodes in the
2 cluster of the cluster IP address and a media access control (MAC) address associated with
3 the node that is bound to the cluster IP address.

1 3. The method of 1 further comprising:

2 (a) forming a virtual local area network (VLAN) comprising:

3 (1) a master node wherein the master node is the node that is bound to
4 the cluster IP address;

5 (2) at least one network interface for each node in the cluster; and

6 (b) wherein multicasting comprises packet forwarding and processing the client
7 request from the master node to the other nodes in the VLAN.

1 4. The method of claim 1 further comprising:
2 forming a multicasting group comprising all of the cluster nodes; and
3 wherein the multicasting comprises automatically multicasting the request to all of
4 the cluster nodes in the multicasting group.

1 5. The method of claim 4 wherein the multicasting group is formed by setting
2 the MAC addresses of network interface cards of nodes in the cluster to be a multicast MAC
3 address.

1 6. The method of claim 5 wherein the MAC addresses are set by setting a first
2 bit of a first byte to a certain value.

1 7. The method of claim 1 wherein the workload distribution function is
2 installed in a driver on each node in the cluster.

1 8. The method of claim 1 wherein the workload distribution function filters the
2 client request based on workload distribution and whether a packet header of the client
3 request indicates that destination MAC and IP addresses are the cluster IP and cluster MAC
4 addresses.

1 9. The method of claim 1 wherein the response is sent from the single node to
2 the client without multicasting.

1 10. The method of claim 1 wherein the workload distribution function
2 distributes the workload by moding a source IP address with a number of nodes in the
3 cluster.

11. The method of claim 1 wherein the workload distribution function distributes the workload by:

representing a total workload observed by the cluster as a bitmap with a number of bits k;

obtaining a bit m by moding a source IP address of the client by the number of bits k; and

assigning the client request to a cluster node that has a specified value at bit m.

12. The method of claim 1 wherein the workload distribution function distributes the workload based on workload statistics that are periodically collected by at least one cluster node.

13. The method of claim 12 wherein cluster nodes periodically exchange workload statistics information.

14. The method of claim 1 further comprising:

adjusting a number of nodes in the cluster;

recomputing a workload distribution based on the number of nodes in the cluster;

and

redistributing the workload among nodes in the cluster based on the recomputation.

15. An apparatus for providing a single system image in a clustered environment comprising:

(a) a master node in a storage cluster, wherein a node in the storage cluster is designated as the master node by assigning an internet protocol (IP) address as a cluster IP

address and binding the cluster IP address to the master node, wherein the master node is configured to:

- (1) receive a client request directed to the cluster IP address;
- (2) multicast the request to all nodes in the cluster;
- (b) at least one additional node in the storage cluster;
- (c) a dynamically adjustable workload distribution function installed on each node in the cluster, wherein the function is configured to filter the client request by allowing a single node to process the client request;

wherein each node in the cluster is configured to:

- obtain a response to the request;
- insert a cluster media access control (MAC) address into the response; and
- send the response from the single node to the client.

16. The apparatus of claim 15 wherein the master node is further configured to inform the other nodes in the cluster of the cluster IP address and a media access control (MAC) address associated with the master node.

17. The apparatus of claim 15 wherein:
the storage cluster comprises a virtual local area network (VLAN); and
the master node is configured to multicast the request by packet forwarding and processing the client request from the master node to the other nodes in the VLAN.

18. The apparatus of claim 15 wherein:
the storage cluster comprises a multicasting group comprising all of the cluster nodes; and

the master node is configured to multicast by automatically multicasting the request to all of the cluster nodes in the multicasting group.

19. The apparatus of claim 18 wherein the multicasting group is formed by setting the MAC addresses of network interface cards of nodes in the cluster to be a multicast MAC address.

20. The apparatus of claim 19 wherein the MAC addresses are set by setting a first bit of a first byte to a certain value.

21. The apparatus of claim 15 wherein the workload distribution function is installed in a driver on each node in the cluster.

22. The apparatus of claim 15 wherein the workload distribution function filters the client request based on workload distribution and whether a packet header of the client request indicates that destination MAC and IP addresses are the cluster IP and cluster MAC addresses.

23. The apparatus of claim 15 wherein the response is sent from the single node to the client without multicasting.

24. The apparatus of claim 15 wherein the workload distribution function distributes the workload by moding a source IP address with a number of nodes in the cluster.

25. The apparatus of claim 15 wherein the workload distribution function distributes the workload by:

representing a total workload observed by the cluster as a bitmap with a number of bits k ;
obtaining a bit m by moding a source IP address of the client by the number of bits k ; and
assigning the client request to a cluster node that has a specified value at bit m .

26. The apparatus of claim 15 wherein the workload distribution function distributes the workload based on workload statistics that are periodically collected by at least one cluster node.

27. The apparatus of claim 26 wherein cluster nodes periodically exchange workload statistics information.

28. The apparatus of claim 15 wherein the workload distribution function is further configured to adjust a number of nodes in the cluster by:
recomputing a workload distribution based on the number of nodes in the cluster;
and
redistributing the workload among nodes in the cluster based on the recomputation.

29. An article of manufacture, embodying logic to perform a method of providing a single system image in a clustered environment, the method comprising:
assigning an internet protocol (IP) address as a cluster IP address;
binding the cluster IP address to a node in a cluster;
receiving a client request directed to the cluster IP address;
multicasting the request to all nodes in the cluster;

7 filtering the request based on a dynamically adjustable workload distribution
8 function, wherein the function is configured to allow a single node to process the client
9 request;
10 obtaining a response to the request;
11 inserting a cluster media access control (MAC) address into the response;
12 sending the response from the single node to the client.

1 30. The article of manufacture of claim 29 wherein the method further
2 comprises informing other nodes in the cluster of the cluster IP address and a media access
3 control (MAC) address associated with the node bound to the cluster IP address.

1 31. The article of manufacture of claim 29, the method further comprising:

2 (a) forming a virtual local area network (VLAN) comprising:

3 (1) a master node wherein the master node is the node that is bound to
4 the cluster IP address;

5 (2) at least one network interface for each node in the cluster; and

6 (b) wherein multicasting comprises packet forwarding and processing the client
7 request from the master node to the other nodes in the VLAN.

1 32. The article of manufacture of claim 29, the method further comprising:

2 forming a multicasting group comprising all of the cluster nodes; and

3 wherein the multicasting comprises automatically multicasting the request to all of
4 the cluster nodes in the multicasting group.

1 33. The article of manufacture of claim 32 wherein the multicasting group is
2 formed by setting the MAC addresses of network interface cards of nodes in the cluster to
3 be a multicast MAC address.

1 34. The article of manufacture of claim 33 wherein the MAC addresses are set by
2 setting a first bit of a first byte to a certain value.

1 35. The article of manufacture of claim 29 wherein the workload distribution
2 function is installed in a driver on each node in the cluster.

1 36. The article of manufacture of claim 29 wherein the workload distribution
2 function filters the client request based on workload distribution and whether a packet
3 header of the client request indicates that destination MAC and IP addresses are the cluster
4 IP and cluster MAC addresses.

1 37. The article of manufacture of claim 29 wherein the response is sent from the
2 single node to the client without multicasting.

1 38. The article of manufacture of claim 29 wherein the workload distribution
2 function distributes the workload by moding a source IP address with a number of nodes in
3 the cluster.

1 39. The article of manufacture of claim 29 wherein the workload distribution
2 function distributes the workload by:
3 representing a total workload observed by the cluster as a bitmap with a number of
4 bits k;

5 obtaining a bit m by moding a source IP address of the client by the number of bits
6 k ; and
7 assigning the client request to a cluster node that has a specified value at bit m .

1 40. The article of manufacture of claim 29 wherein the workload distribution
2 function distributes the workload based on workload statistics that are periodically collected
3 by at least one cluster node.

1 41. The article of manufacture of claim 40 wherein cluster nodes periodically
2 exchange workload statistics information.

1 42. The article of manufacture of claim 29, the method further comprising:
2 adjusting a number of nodes in the cluster;
3 recomputing a workload distribution based on the number of nodes in the cluster;
4 and
5 redistributing the workload among nodes in the cluster based on the recomputation.

Approved for Release by NSA on 09-08-2013 pursuant to E.O. 13526